MOLECULAR AND CELLULAR BIOLOGY, Nov. 1995, p. 6291–6298 Vol. 15, No. 11
0270-7306/95/$04.00+0

Selection of Novel Exon Recognition Elements from a Pool of Random Sequences

HUICHENG TIAN AND RYSZARD KOLE*

Lineberger Comprehensive Cancer Center, Department of Pharmacology, and Curriculum in Genetics, University of North Carolina, Chapel Hill, North Carolina 27599

Received 9 May 1995/Returned for modification 21 June 1995/Accepted 28 July 1995

A 20-nucleotide sequence close to the 3′ end of the internal exon of a model two-intron, three-exon pre-mRNA (DUP184 [Z. Dominski and R. Kole, J. Biol. Chem. 269:23590–23596, 1994]) was replaced by a random 20-mer, resulting in a pool of pre-mRNAs which, like the initial DUP184 construct, were spliced in vitro by a pathway leading to predominant skipping of the internal exon. The randomized pre-mRNAs were subjected to a selection protocol, resulting in a pool enriched in pre-mRNAs that efficiently included the internal exon. Isolation and sequencing of a number of clones corresponding to the selected pre-mRNAs showed that two classes of sequences were selected from the initial pool. Most abundant among these were sequences with purine tracts similar to those in the recently identified exon-splicing enhancers, while a smaller class included sequences lacking discernible purine tracts within the 20-nucleotide region. Splicing of selected pre-mRNAs showed that the purine tracts vary in their ability to promote exon inclusion and, more important, that sequences lacking purine tracts stimulate inclusion of the internal exon as efficiently as their purine-rich counterparts. The latter result indicates the existence of a novel class of exon recognition sequences or splicing enhancers.

The role of exons and exon sequences in pre-mRNA splicing and splice site selection was first observed several years ago (14, 15, 31, 34). The early work as well as subsequent reports showed that short exons greatly decrease the efficiency of in vitro splicing in monointronic pre-mRNAs (14, 15, 30, 31, 34, 47) and are frequently skipped from multiintronic pre-mRNAs spliced both in vitro and in vivo (2, 3, 10, 11, 28, 36). The experimental data were in agreement with the statistical analysis of exon length, which showed that an average vertebrate internal exon is 137 nucleotides long and that internal exons shorter than 50 nucleotides are present in less than 5% of the analyzed sequences (18). Recent results showed that the upper limit on exon length is not as stringent, and internal exons of up to 1,500 nucleotides have been shown to be efficiently included in the spliced RNA (5). A possible mechanism for skipping of short exons appears to be an interplay of steric hindrance between the splicing factors bound to splice sites flanking the exon and their affinities for these sites (2, 3, 11, 25, 28, 36, 40). The cumulative evidence led to a proposal (32), supported by subsequent observations (11, 25, 36, 40), that exons and not introns constitute the element in pre-mRNA that is initially recognized by *trans*-acting splicing factors.

Specific exon sequences were found to be essential for alternative splicing of a cascade of gene products that determine gender-specific development in *Drosophila* flies (19, 21, 29, 42, 43, 44), for proper splicing in certain retroviruses (13, 22), and for tissue-specific splicing in a number of other pre-mRNAs (8, 17, 37, 45, 49). In some instances, even one or two point mutations within an exon sequence completely abolished its recognition by the splicing machinery (6, 7). Recently, polypurine tracts were identified in a number of alternatively spliced exons and found to promote splicing at the upstream 3′ splice site (9, 26, 38, 41, 50–52). At least in some pre-mRNAs, these sequences appear to have enhancer-like properties, since they increase internal exon inclusion from various positions within the exon (26, 41). There appears to be a range of purine sequences that stimulate splicing, although poly(A), poly(G), and purines interrupted by uridines are not effective (41). The latter observation is in agreement with our early work, which showed that pyrimidine tracts (mostly uridines) act as splicing "poison sequences" when inserted in the downstream exon of monointronic pre-mRNA (15). Although purine-rich splicing enhancers clearly stimulate splicing in some pre-mRNAs, it seems unlikely that they are the only exon sequences responsible for recognition of internal exons. Moreover, polypurine sequences may not be effective in all sequence contexts. For example, a fragment of the human immunodeficiency virus *tat* exon inhibits splicing at the upstream 3′ splice site in spite of the fact that it contains a purine-rich tract (1). Thus, it seems possible that exons contain both negative and positive splicing elements which play a role in determining the balance between exon inclusion and skipping.

Identification of functional exon sequences by mutations of alternatively spliced exons, although useful, limits one to a relatively small number of exons. Furthermore, the types of introduced mutations are usually based on preconceived notions on the role of specific sequence elements in splicing; hence, this approach may easily miss important sequences that are not obvious at the outset.

In this study, by taking advantage of molecular selection techniques (16, 20a, 46), exon sequences that stimulate inclusion of the internal exon into the mRNA have been selected from a pool of approximately 10^{12} sequence species. As a model, we have used three-exon pre-mRNAs, in which exon inclusion patterns are easily detectable. Splicing of these pre-mRNAs had previously allowed us to demonstrate that the lengths of exons (10), their interactions with other sequence elements in pre-mRNA (11), and their sequences (12) affect exon recognition and splice site selection, both in vitro and in vivo. The present findings indicate that several of the selected

* Corresponding author. Mailing address: University of North Carolina, Lineberger Comprehensive Cancer Center, CB#7295, Chapel Hill, NC 27599. Phone: (919) 966-1143. Fax: (919) 966-3015.

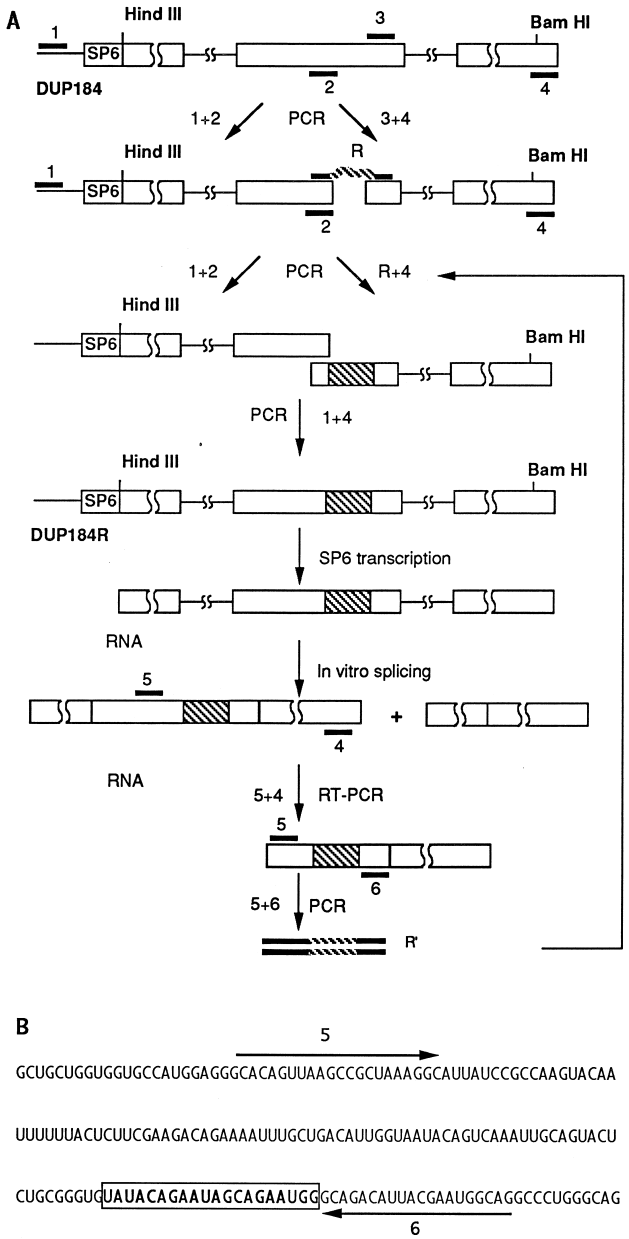

FIG. 1. (A) Selection procedure. Boxes, exon sequences; thin lines, introns; SP6 box, SP6 promoter; heavy lines, amplimers 1 to 5 and R. Amplimer R contains the 20-nucleotide random sequence (stippled). Incorporation of this sequence by PCR, as shown, into the internal exon of DUP184 results in the DUP184R construct, which represents a pool of DNA templates containing the random sequence. Transcription and splicing of DUP184 leads to predominant internal exon skipping, but that of DUP184R yields spliced RNA, a small fraction of which efficiently includes the internal exon. This RNA fraction is amplified by RT-PCR, and the product (R′) is isolated. The RT-PCR product is enriched in sequences that promote internal exon inclusion and is used in the next cycle of PCR, transcription, splicing, and RT-PCR (see Materials and Methods and Results for details). (B) Sequence of the internal exon of DUP184 pre-mRNA. The randomized sequence is boxed; arrows indicate amplimers 5 and 6, respectively.

sequences contain polypurine tracts, as could be expected in the light of previous results (9, 12, 26, 27, 35, 38, 41, 51, 52), and that the purine-rich sequences vary in their ability to promote exon inclusion. More important, we have found sequences that lack polypurine tracts yet efficiently stimulate inclusion of the internal exon, which points to the existence of a novel group of exon recognition sequences or splicing enhancers.

MATERIALS AND METHODS

The DUP184 construct that served as the starting material for the selection procedure was described previously (12). All amplimers used in the selection procedure are illustrated schematically in Fig. 1A. Oligonucleotide R, used to generate the randomized sequence pool, included 18 and 20 nucleotides of flanking sequence upstream and downstream, respectively, of the random 20-mer. The sequences of amplimers 5 and 6 are shown in Fig. 1B. The sequences of amplimers 1 to 4 are: amplimer 1, TCTGCTTCAGTAAGCCAGATGCTAC ACAAT; amplimer 2, CACCCGCAGAGTACTGCAATTTGACT; amplimer 3, GCAGACATTACGAATGGCAG; and amplimer 4, CCTGAAGCTTTCAGG ATCCCACGTGCAGCT. The sequences of the internal exon and the randomized fragment are shown in Fig. 1B.

Transcription of DNA templates by SP6 RNA polymerase, preparation of nuclear extracts from HeLa cells, and in vitro splicing were done by established procedures (23, 24) as described previously (11). The DNA templates used in the experiments shown in Fig. 4 were PCR products, representing pools from each selection cycle. The individual plasmids that served as transcription templates for the splicing experiments shown in Fig. 3 and 8 were prepared by the boiling miniprep method; those shown in Fig. 6 were purified through CsCl gradients (33). The plasmids were linearized with *Bam*HI before transcription. After in vitro splicing, the RNA used in the selection procedure (see below) was extracted with Tri-Reagent (Molecular Research Center, Cincinnati, Ohio) to remove the DNA template.

The molecular selection procedures (16, 20a, 46) were carried out as schematically represented in Fig. 1A and discussed in the text. r*Tth* thermostable DNA polymerase was used for the reverse transcription step, and either r*Tth* or *Taq* polymerase was used for the PCR step of the selection procedure. The reaction conditions were as recommended by the manufacturer (Perkin Elmer Cetus). The spliced RNA isolated from 80 μl of in vitro splicing reaction mix was subjected to reverse transcription in a 20-μl and subsequent PCR in a 100-μl reaction volume. The primers were used at 0.6 μM. Although the random 20-mer oligonucleotide represents a pool of approximately 10^{12} different sequences, the reaction conditions may not be sufficient to reach this number.

All autoradiograms were captured with a DAGE MTI CCD72 video camera (Michigan City, Ind.), and the images were processed with NIH Image 1.47 and MacDraw Pro 1.0 software. The efficiency of internal exon inclusion was quantitated by densitometry of the autoradiograms shown in Fig. 3, 6, and 8. The results are expressed as a percentage of the product with included internal exon relative to the sum of the products with the exon either skipped or included. The final figures were printed out on a Sony dye sublimation printer.

RESULTS

We have previously obtained two-intron, three-exon constructs, based on the human β-globin pre-mRNA, in which the internal exons of approximately the same length were either skipped or included, depending on their sequence; these constructs provided an excellent model system for studies on exon recognition sequences (12). It was found that a 184-nucleotide-long internal exon (DUP184), containing a sequence from the yeast *URA3* gene flanked by human β-globin splice sites, was skipped during splicing in HeLa nuclear extract and in transfected HeLa cells (see Fig. 1 for the structure of the construct). Skipping was partially reversed if the DUP184 sequence was mutated to include a purine-rich tract, an oligo(C) tract, or a fragment of the human β-globin exon (12) (see also the DUP184AG construct [Fig. 3, 5 and 7]). To gain a more general view of exon recognition sequences, we developed a PCR-based in vitro evolution protocol (16, 46) for selection of sequences that prevent skipping of the DUP184 exon and stimulate its inclusion.

Analysis of randomized constructs. The scheme of the randomization and selection procedure is shown in Fig. 1A. First, the 5′ and 3′ fragments of the DUP184 construct were obtained by PCR with amplimers 1 and 2 and amplimers 3 and 4, respectively. To introduce the random sequence, the 3′ fragment of DUP184 was subjected to PCR with amplimers R and 4; the former is a 58-mer oligonucleotide containing 20 nucleotides of random sequence and 18- and 20-nucleotide flanks

complementary to the 3′ portion of the DUP184 internal exon. The two-step procedure was designed to eliminate selective hybridization of some species of amplimer R to the DUP184 sequence. The 5′ half and the randomized 3′ half, which now overlap by 18 nucleotides (overlap of amplimers R and 2), were combined, annealed, and subjected to PCR with amplimers 1 and 4. This procedure generated a pool of DNA templates (DUP184R [Fig. 1A]) containing the SP6 promoter and the full-length equivalent of the DUP184 sequence containing the randomized sequences in the 3′ portion of the internal exon.

To gain insight into the sequences of the randomized pool, the *Hin*dIII-*Bam*HI fragment of the PCR-generated templates (Fig. 1A) was subcloned into the SP64 vector, and 30 individual clones were isolated and sequenced as described under Materials and Methods. The replacement of DUP184 sequences was clearly successful, since all but one of the constructs showed modified sequences within the targeted region (Fig. 2). The single unmodified clone (clone 0.30 [Fig. 2]) probably originated from the DUP184 plasmid used as the initial template for PCR. Of the sequenced clones, 19 contain the full-length 20-mer, while 11 clones are truncated within the randomized sequence. These deletions are most likely due to hybridization during PCR of the overlapping region of the 5′ half to sequences within the randomized 20-mer. The pool of 20-mer sequences has an apparent purine bias (the average purine content of the full length 20-mer is 67%, ranging from 45 to 85%). Moreover, the first nucleotide of the 20-mer is an adenosine in 10 of the 19 full-length clones and in 19 of all 30 clones. A high frequency of adenosine at the first position was also detected by sequencing the total pool of DNA templates (data not shown). These deviations from random sequence were generated either during the chemical synthesis of amplimer R or by PCR.

To test the splicing characteristics of the modified constructs, 10 clones were transcribed by SP6 RNA polymerase, and the pre-mRNAs were spliced in a HeLa nuclear extract. They included five consecutively isolated full-length constructs (clones 0.2, 0.3, 0.4, 0.6, and 0.7) and those that contained short purine tracts (defined as four or more consecutive purines) at the 5′ end (0.12 and 0.17), at the 3′ end (0.13), or in the middle (0.15) of the 20-mer sequence (Fig. 2). Pre-mRNA 0.19 (Fig. 2), with the 85% purine-rich 20-mer sequence, was also tested in the splicing assay.

Splicing of the pre-mRNA transcribed from the DUP184 construct, the starting material for the selection procedure, led to predominant skipping of the internal exon (Fig. 3, lane 2), while splicing of the control DUP184AG pre-mRNA (Fig. 3, lane 1), containing a polypurine tract in the internal exon, resulted in a marked increase in exon inclusion, as previously observed (12). With the exception of constructs 0.6 and 0.7 (Fig. 3, lanes 6 and 7, respectively), all other constructs included the internal exon either as inefficiently as the starting DUP184 pre-mRNA (compare lane 2 with lanes 5, 8, 9, and 19 in Fig. 3) or only slightly better (Fig. 2, lanes 3, 4, 10, and 11). These results lead to several conclusions.

First, the fact that construct 0.7, which is efficient in exon inclusion (Fig. 3, lane 7), was found among the 10 unselected clones predicts that it will be possible to select similar constructs from the original pool of randomized PCR products. Second, the replacement of the DUP184 sequence with modified sequences that do not promote exon inclusion indicates that the exon inclusion seen in construct 0.7 and other selected clones (see below) is caused by positively acting sequences introduced in the randomized fragment and not by removal of a negative element from the DUP184 RNA (see also Discussion). Third, the enhancement of exon inclusion does not correlate well with the purine content or the length of the purine tracts within the 20-mer sequence. The efficiently included exon in pre-mRNA 0.7 (Fig. 2 and Fig. 3, lane 7) has 55% purines and a single six-nucleotide uninterrupted purine tract within the randomized 20-mer, whereas pre-mRNA 0.19 (Fig. 3, lane 12) and, for that matter, DUP184 (Fig. 3, lane 2), with purine contents of 85 and 70% and three and two purine tracts, respectively, are unable to include the internal exon. These observations suggest either that specific purine tract sequences may act as splicing enhancers or that other sequences within the randomized fragment are equally capable of promoting exon inclusion.

CYCLE 0

184	GUACUCUGCGGGUG UAUACAGAAUAGCAGAAUGG GCAGACAUU
0.2	G CAUUGAGGGAGAUAUCAGGA G
0.3	G CUCUGACAGCGUCAAAGUUG G
0.4	G UUCGAGAGCAGUUAACAGUG G
0.6	G ACAAAUCGUAUGCGGAGAUU G
0.7	G AUUUCGGCGCCUGGAGAACA G
0.10	G GCAAAUAAGGGAAGACGCGA G
0.12	G ACGUUUAGUACCACGAAAGU G
0.13	G AGAGGAUAGUUGUAGAUACG G
0.14	G AGCCGAAAGCAGUGAAGAGC G
0.15	G CCUGCUCUAAGACAACCUAA G
0.16	G AGGAAAAUUAUCGCACUUGA G
0.17	G UGGACAGCACCAGCAAGGGA G
0.19	G AGACAGGGCAGAAAGGGAUG G
0.20	G GAGAAACGGAUAAGAAGAUC G
0.25	G GCGAAGGCGAAACAUGGCGA G
0.26	G AACGCAAGGACGUGGUAAGG G
0.27	G AGGAGUCAACAGGUGGCGGG G
0.29	G AAACAGACGCCGCACGAGUG G
0.30	G UAUACAGAAUAGCAGAAUGG G
0.1	G AAAACCGAAUGCUAACAU-- G
0.5	G AUUAAGAGGGGUUACGG--- G
0.8	G AGAGCGGAAGGCAAAGUGU- G
0.9	G AAUAAGGGACGCGCAAG--- G
0.11	G AAACAUCAAUACCGGAAGG- G
0.18	G GGAUAUAACGACAGUAAAG- G
0.21	G GACGUAGAGUUCUUUAG--- G
0.22	G ACCAGAAUAGUCUAGAGCG- G
0.23	G AGGGAACA---------- G
0.24	G AUCGUCGAAAAGGAGA---- G
0.28	G AGGGAGGGAGGUUC------ G

FIG. 2. Sequences of the randomized region in individual clones obtained from the initial pool of DNA templates. The sequence of DUP184, the starting material, is at the top. The lower part of the figure shows the sequences of the clones that have deletions within the randomized 20-mer. Tracts of four or more consecutive purines are shown in boldface.

Selection of exon recognition sequences. To select constructs that contain exon recognition (enhancer) sequences, the pre-mRNA was transcribed from the total pool of DNA templates containing the random sequences (DUP184R) and spliced in a HeLa extract. Total RNA from the splicing reaction mix was subjected to reverse transcription-PCR (RT-PCR) with amplimers 4 and 5, the amplified products were resolved on an

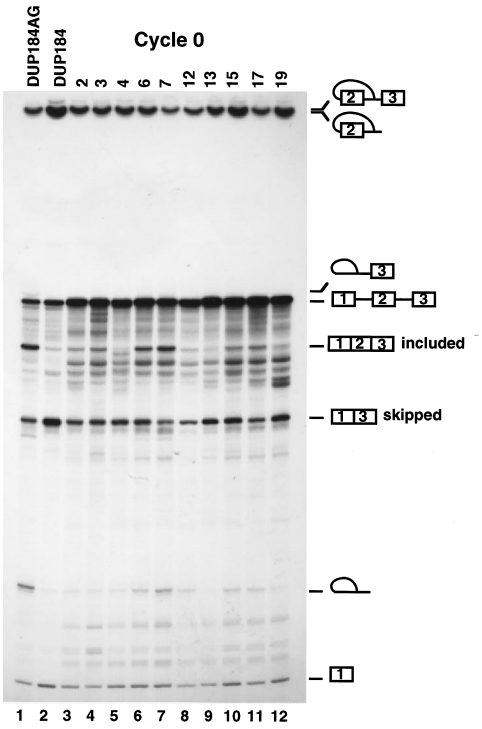

FIG. 3. Splicing of individual pre-mRNAs transcribed from the constructs shown in Fig. 2. Splicing of control pre-mRNAs which include (DUP184AG) or exclude (DUP184) the internal exon (12) is shown in lanes 1 and 2, respectively. Lanes 3 to 12, splicing of pre-mRNAs transcribed from the individual clones indicated at the top of the figure. The structures of the products and intermediates of splicing are shown on the right. Some of the bands migrating between the included and skipped spliced products represent additional splicing intermediates (see Fig. 6), while the others are contaminants arising from transcription of the pre-mRNA from the crude, miniprep DNA templates (see Materials and Methods). This procedure was used in this experiment and the one shown in Fig. 8.

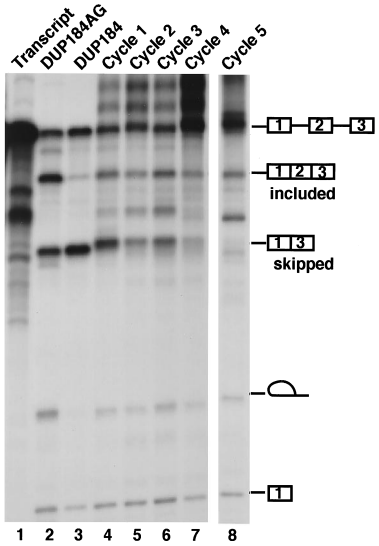

FIG. 4. Splicing of selected pools of pre-mRNAs. The RNA from DUP184R pooled templates was transcribed and spliced in a HeLa cell extract. Lane 1, unspliced transcript from cycle 3 of the selection procedure; lanes 2 to 8, splicing of the following pre-mRNAs: lane 2, DUP184/AG; lane 3, DUP184; lanes 4 to 8, pre-mRNAs from cycles 1 to 5, respectively, as indicated at the top of the figure. The structures of the products and intermediates of splicing are shown on the right. The band migrating on the gel between the skipped and included spliced products is not generated during splicing but originates as a result of pretermination or degradation during SP6 RNA polymerase transcription of the pooled PCR-derived templates.

agarose gel, and the band corresponding to the spliced three-exon mRNA was excised and eluted. The last step removed RT-PCR products that may have been generated on unspliced pre-mRNA present in the splicing mixture. In the subsequent step, the randomized sequence of the internal exon, which in principle could have originated only from a pool of sequences that enhance exon inclusion, was amplified with amplimers 5 and 6, yielding product R′. The R′ fragment was used as an amplimer in the next round of PCRs to generate first the overlapping halves and then the full-length DUP184R′ template. In this step, the random sequence, now presumably enriched in exon inclusion enhancers, was inserted back into the DNA template containing the introns. This material was used in the next cycle of transcription, splicing, and selection, as shown in Fig. 1A.

Figure 4 shows the analysis of in vitro splicing of the pooled pre-mRNA mixtures subjected to five cycles of splicing and selection. Splicing of the pre-mRNA pooled from the first selection cycle already showed enhancement of exon inclusion relative to DUP184 pre-mRNA, as indicated by the increased amount of the spliced product containing all the three exons (Fig. 4, lane 4). Exon inclusion increased further with subsequent cycles (Fig. 4, lanes 5 to 7), so that after cycle 5 (lane 8), most of the pooled pre-mRNA was spliced via the exon inclusion pathway.

Analysis of selected constructs. After cycle 5, the amplified material (Fig. 1A) was subcloned, and the individual clones were isolated and sequenced. We have obtained sequences of the 22 individual clones that contained the selected full-length 20-mer (Fig. 5, top) and, in addition, of the 10 clones that were truncated within the 20-mer sequence (Fig. 5, bottom).

As expected, the differences in sequence were localized to the randomized 20-mer, and no additional mutations, possibly introduced by the multiple rounds of PCR, were detected in the sequences adjacent to this region. Among the 22 clones containing the full-length 20-mer sequence, approximately 70% of the selected sequences begin with A and end with U (Fig. 5), a significant change in comparison to the unselected 20-mers, of which 50% start with A and 11% end with U (Fig. 2). There is no indication of a clear consensus at the remaining positions of the selected sequences. Like the unselected clones, those from cycle 5 retain the high purine content, averaging 63% and ranging from 50 to 75%, and most of them contain purine tracts. The nature of the tracts seems to be similar in the full-length and truncated clones, suggesting that deletions do not significantly affect the underlying mechanism of selection.

Since purine tracts were identified as exon-splicing enhancers (9, 12, 26, 27, 35, 41, 51, 52), it seemed likely that they are responsible for the increase in exon inclusion seen in splicing of the cycle 5 RNA pool. It was then surprising that several clones which, in spite of their relatively high purine content, lacked clearly discernible purine tracts which may be considered splicing enhancers (Fig. 5, clones, 5.3, 5.13, 5.22, and 5.33) were selected. To investigate whether variations in the sequence of the selected clones and in particular the lack of polypurine tracts affect the efficiency of exon inclusion, several clones were transcribed, and their RNA was spliced in vitro in nuclear extracts from HeLa cells.

Constructs with single or multiple purine tracts within the randomized sequence and especially those lacking tracts of four consecutive purines were chosen for splicing. To eliminate

CYCLE 5

184	GUACUCUGCGGGUG	UAUACAGAAUAGCAGAAUGG	GCAGACAUU
5.1	G	ACGGUCUUCAAUGAGGCGAU	G
5.2	G	UGAGAAGGAACGGCCACAAU	G
5.3	G	AUGAUCAUCGGAUGACAUAG	G
5.4	G	AUGGAACAAAGAUGAGAACU	G
5.7	G	AUAUCAAGAAGCAGAGACGU	G
5.9	G	UACGGAGACUUUAACCAUGU	G
5.13	G	CCCGGAUAUUGAUACGAGCA	G
5.14	G	ACAUCAUCUACCGAAGAUGU	G
5.16	G	AGGACGAUCAUCGGACCGAU	G
5.17	G	AGAGAUAUCACGGGACGACC	G
5.21	G	ACGACCGGAAACAAGUGUGU	G
5.22	G	CAAUGGACAUAUAACGGCUG	G
5.25	G	CGAGGACGUAUUGAACGACA	G
5.26	G	AAACGUAAAGUACGCCGUUU	G
5.27	G	ACGGCACGUGGACGAAGUGG	G
5.30	G	AGGAAGCGAGCCAGUUGGAU	G
5.32	G	AAACUACAAGAAUUCUGGGU	G
5.33	G	UGACGGUGAACGCUGGGUAU	G
5.35	G	AAGUUGACGUGAACGACGGU	G
5.39	G	ACGUGGAGAGGGUGAUGAGU	G
5.42	G	UGCCACAUCGGAGACGACGU	G
5.44	G	AGCACGCAUGAAAACCUGGU	G
5.8	G	AACGGAGGAAGCAGCUGGA-	G
5.10	G	GUCCGGAUUGGGAAGU----	G
5.15	G	AGGACAGAUACGUCAGACU-	G
5.20	G	CGCCAAGCGUA---------	G
5.24	G	AUAUAGCAGACCACGAAGG-	G
5.31	G	CGUGUC----------	-
5.34	G	CAUGCGAAGAAGGAU-----	G
5.38	G	AAUCGUGCCGG---------	G
5.41	G	GAUCGAGAACAACACGUGG-	G
5.43	G	AAACACAUGCUGGA------	G

FIG. 5. Sequence of the randomized region from individual clones obtained after the fifth cycle of selection. Details are as described in the legend to Fig. 2.

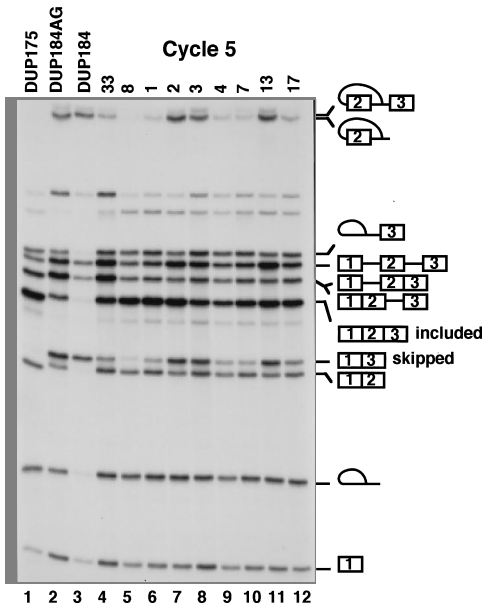

FIG. 6. Splicing of individual pre-mRNAs selected in cycle 5. Lanes 1 to 3, splicing of control pre-mRNAs DUP175, DUP184AG, and DUP184 (12), respectively; lanes 4 to 12, splicing of pre-mRNAs transcribed from the individual clones indicated at the top of the figure. The structures of the products and intermediates of splicing are shown on the right. The DNA templates used for pre-mRNA transcription were purified by CsCl gradient centrifugation.

a possible effect of internal exon length on the splicing pathway, splicing was carried out with pre-mRNAs containing the full-length 20-mer sequence. Consistent with the results of splicing of pooled RNA from cycle 5 (Fig. 4, lane 8), all pre-mRNAs tested included an internal exon much more efficiently than the starting construct, DUP184 (Fig. 6, lane 3). However, as has already been noticed for unselected clones, the efficiency of exon inclusion did not correlate well with the length or the "strength" of the polypurine tracts. For example, clones 5.4, 5.7, and 5.17 (Fig. 6, lanes 9, 10, and 12, respectively), which contain more than one polypurine tract (see Fig. 5), were not as efficient in exon inclusion as clone 5.8 (Fig. 6, lane 5), in which the 20-mer sequence is truncated by one nucleotide and contains a single stretch of eight consecutive purines. In fact, the three former clones included the exon about as well as clone 5.1 (Fig. 6, lane 6), which contains only a single purine tetramer and, in addition, four uridines, nucleosides reported to interfere with exon enhancer activity (41). The splicing of clone 5.33 (lane 4) was even more striking. This clone has no polypurine sequence yet was about as efficient in exon inclusion as the clones containing the purine tracts. In fact, it was somewhat more efficient in exon inclusion than clone 5.2 (Fig. 6, lane 7), with a 9-mer sequence, GAGAAGGAA (Fig. 5), which conforms to a presumed consensus sequence for purine splicing enhancers (27, 41). Among the clones lacking the polypurine tracts, the ability to promote exon inclusion was not unique to clone 5.33, since two other constructs from this group, clones 5.3 and 5.13, included the internal exon with an efficiency similar to that of clone 5.2 (Fig. 6, lanes 8 and 11, respectively). These results clearly indicate that certain sequences lacking purine tracts serve as exon recognition signals about as efficiently as polypurine sequences.

Further selection of exon recognition sequences. Since there was an apparent lack of consensus in the selected 20-mer sequences and variability in the efficiency of exon inclusion among the pre-mRNAs selected through cycle 5, we subjected the selected material to two additional selection cycles, with subsequent cloning and analysis along the lines discussed above. We expected that additional cycles might eliminate the less efficiently included exon sequences and narrow down the spectrum of selected sequences.

Sequence analysis of the 26 full-length clones (Fig. 7) shows that there is only a small, if any, drift in the overall character of the selected sequences compared with those after five cycles. The first adenosine and the last uridine of the 20-mer sequence are now present in over 80% of the clones. There is still a lack of consensus at other positions, and the average purine content (64%) is not significantly changed from that of cycles 0 and 5. Several clones lack polypurine tracts; these include three clones with intact 20-mer sequences and six with truncated sequences (not shown). The average uridine content is reduced from four to three residues per 20-mer sequence, while the average number of cytidines is increased slightly, if at all (19 to 21%). There are only two significant differences due to the additional two cycles. First, the number of truncated sequences increased from one-third to almost half of the selected clones (not shown). Second, in the majority of clones, the flanking

CYCLE 7

184	GUACUCUGCGGGUG	UAUACAGAAUAGCAGAAUGG	GCAGACAUU
7.1	G	GCCUCGCGAAAAGGAGACGU	G
7.3	G	ACGUGAAACGGCACGUAUCU	G
7.4	G	AUACGGAAAACAUGACGACU	G
7.5	G	AAGGGAUGAACGACUGGACU	G
7.6	G	ACGAGGUCGGGAGCCUGUGU	G
7.7	G	GAAGGAAAUGUGCGACGGCU	G
7.8	G	AAACAACGGUGUCGAACCGU	G
7.10	G	GAGCAAGGAGCAACGGCGAU	G
7.11	G	AAAGAAGUCGCGGAACGGAU	G
7.16	G	AAGAAGGCAAACAUGAGGAU	G
7.19	G	AGAACGGAGGACGCCGGCGU	G
7.21	G	ACAUAGAAAUGAAGAUUGCU	G
7.23	G	AUGGAGUCGUCGGAGGACUU	G
7.24	G	AAACAAACCAGAGGAUCGCA	G
7.25	G	AUCCGUUUGAUGUGGAUGGA	G
7.26	G	AAUCAGCGAACAAGCCAUCU	G
7.29	G	AUACAGACAACGGUUCAACG	G
7.30	G	AAGAUGUUCAUCGUGAGGCU	G
7.31	G	ACAGCAGCGACGAAGAACUU	G
7.34	G	AGGAACCAGACCUUGGGUGC	G
7.37	G	GAGACAACCGAUCGAACAGU	G
7.39	G	AACUACAGACGGCCGAACGU	G
7.46	G	ACGGCCUACGCGACAAGGGU	G
7.48	G	AGAGCCAGAACAGCGCAACG	G
7.49	G	GAACGAGCGAACAUCGGUGU	G
7.50	G	GAUCGAAGGAAGAAGUCGCU	G
7.14	G	AUCGGGCAACCAGU------	G
7.32	G	AUCUUAGCAUGCGACUGGC-	G
7.36	G	AAUCUGAUCACU--------	G

FIG. 7. Sequences of the randomized region from individual clones obtained after the seventh cycle of selection. Details are as described in the legend to Fig. 2.

sequence immediately adjacent to the 3′ end of the randomized 20-mer contained two to four repeats corresponding to the sequence of amplimer 6 (data not shown). This is probably because part of this sequence, at the 3′ end of amplimer 6, GCAG, is repeated three times in the downstream region of the exon (see Fig. 1B). It is not clear why these aberrant PCR products became dominant after the seventh cycle of the procedure, since the repeats do not appear to promote more efficient exon inclusion (see below).

The efficiency of exon inclusion was tested with five pre-mRNAs containing polypurine tracts (Fig. 7, clones 7.1, 7.11, 7.21, 7.23, and 7.25) and four pre-mRNAs lacking them (Fig. 7, clones 7.14, 7.29, 7.32, and 7.36). In the latter group, clone 7.29 had the full-length 20-mer sequence, clone 7.32 had a 1-nucleotide deletion, and clones 7.14 and 7.36 had 6- and 8-nucleotide deletions, respectively. As with the constructs selected after five cycles, the observed variability in the efficiency of exon inclusion for cycle 7 pre-mRNAs did not correlate well with the purine tract content of the 20-mer sequence. The internal exon of clone 7.11, with two purine tracts of 8 and 4 nucleotides, was included most efficiently (Fig. 8, lane 2). The remaining constructs containing purine tracts included the internal exon with an efficiency varying from 57 to 84% (Fig. 8, lanes 1 and 4 to 6). A similar efficiency of exon inclusion was observed for clones 7.29, 7.32, and 7.36, which are devoid of polypurine sequences (Fig. 8, lanes 7 to 9, respectively). Clone 7.14, with a truncated 20-mer sequence lacking purine tracts (Fig. 8, lane 3), included the internal exon least efficiently but still to a much greater extent than the initial DUP184 pre-mRNA. The exon sequence downstream from the selected sequence containing the amplimer 6 repeats does not appear to affect the course of splicing. This is best seen by comparison of clones 7.1, 7.11, and 7.21, which have the same four repeats of the primer sequence (data not shown) but differed in efficiency of exon inclusion. Overall, the results indicate that in spite of the additional selection steps, the variability in the efficiency of exon inclusion was maintained, as was exon inclusion in the absence of polypurine tracts.

DISCUSSION

By using a PCR-based molecular selection procedure, we have identified a number of exon sequences that promote exon inclusion. This approach provides a powerful method for identification and functional testing of exon recognition elements without preconceived notions as to their sequence or role in splicing. Furthermore, the sequences studied are not limited to those originating from previously identified alternatively spliced exons, and their function can be determined in any sequence context and in conjunction with different splice sites, as exemplified by the model constructs based on the human β-globin gene investigated in this study.

From the initial pool of up to 10^{12} sequences, the procedure rapidly selected those that promoted inclusion of internal exons in spliced RNA, as evidenced by a marked increase in the three-exon spliced product after a single selection step and nearly complete internal exon inclusion after cycle 5 of the procedure (Fig. 4). Analysis of the individual constructs showed that the selection generated a surprisingly wide spectrum of 20-mer sequences, all of which markedly enhanced exon inclusion in comparison to the initial DUP184 construct. Note that after cycles 5 and 7, we have chosen for in vitro splicing a number of constructs that appeared to lack typical purine enhancers. This may be responsible for the fact that the efficiency of exon inclusion in the total pool of cycle 5 (Fig. 4, lane 8) appears to be higher than that of the individual clones selected in this cycle. The enhanced exon inclusion is due to positive effects of the introduced sequences rather than the removal of negative elements from the DUP184 internal exon. This is seen by the facts that only 1 out of 10 sequences stimulated efficient exon inclusion in the unselected pool (Fig. 3), whereas a number of such sequences were identified after several cycles of selection, and that no negative elements could be identified in DUP184 pre-mRNA by extensive mutational analysis of the internal exon (12).

The synthesis of the random oligonucleotide and the PCR resulted in an initial DNA template pool that was biased towards a high purine content in the randomized sequence (67%). Interestingly, selection, even from such a biased pool, led to identification of several constructs that lacked polypurine tracts yet promoted exon inclusion as efficiently as those with previously identified purine-rich splicing enhancer sequences (9, 26, 27, 35, 38, 41, 51, 52). These constructs clearly represent a novel group of exon recognition sequences, distinct from the purine motifs and from an oligo(C) element identified previously (12). The non-purine tract sequences were found in 8 of the 48 full-length clones and in 9 of the 30 clones with a truncated 20-mer sequence. Thus, although less frequent than the polypurine tracts, they represent a significant fraction of the total sample. Like the total sequence sample,

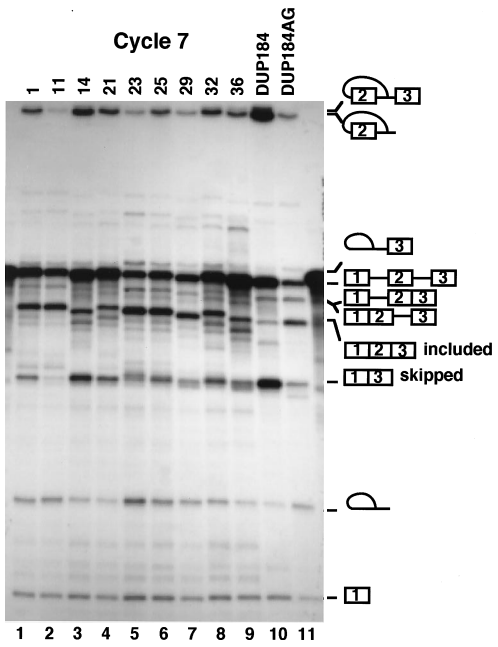

FIG. 8. Splicing of individual pre-mRNAs selected in cycle 7. Lanes 10 and 11, splicing of control pre-mRNAs DUP184 and DUP184AG, respectively; lanes 1 to 9, splicing of pre-mRNAs transcribed from the individual clones indicated at the top of the figure. The structures of the products and intermediates of splicing are shown on the right. The pre-mRNAs used in this experiment were transcribed from crude DNA obtained by the boiling miniprep procedure, most likely causing the additional bands visible on the gel.

clones lacking polypurine tracts do not exhibit a consensus for most of the positions in the aligned full length 20-mer sequence, indicating that a number of different sequences may promote exon inclusion. Such a diverse group of sequences would certainly remain undetected without the powerful selection procedure employed in this work. Note, for example, that a possible member of this group was generated only accidentally in a control experiment in a recent study (35).

The selection procedure does not seem to affect the overall purine content or the nature of the purine tracts. The only apparent difference between the unselected and selected clones is a large increase in the frequency of constructs with randomized sequences starting with adenosine and ending with uridine nucleotides. The significance of this observation is unclear. One possibility is that these nucleosides contribute to a common secondary or tertiary structure that is required for efficient exon recognition. Alternatively, adenosine and uridine create potential pseudo-splice sites flanking the 20-mer sequences (GGUGANNN . . . NNNUGCAG versus G/GUAA GU . . . YNCAG/G for consensus splice sites, where Y is U or C). Although we have excluded the direct positive role of cryptic splice sites in stimulating exon inclusion in the DUP184 construct (12), these structures may contribute to the selection mechanism.

Most of the previous studies on exon splicing enhancers tested for improvement of splicing efficiency of single-intron pre-mRNAs (9, 26, 38, 41, 50–52). In these studies, the activity of the enhancer did not appear to rigidly depend on the sequence of the uninterrupted purine tracts, although significant modifications, such as insertions of a large number of pyrimidines, uridines in particular, or replacement with purine homopolymers (41), abolished their activity. The assay used in this study, based on the analysis of splice site selection as reflected in skipping or inclusion of the internal exon, shows that the sequence of the polypurine tract and apparently the context within the 20-mer sequence have a significant effect on the efficiency of exon inclusion. For example, polypurine tracts with the sequence GAAGGA, representing a consensus exon splicing enhancer (27), in some instances promote exon inclusion to a lesser extent than other sequences, including those lacking polypurine tracts. Nonetheless, the frequency with which polypurine tract sequences are selected in our procedure suggests that they represent a class of important exon recognition sequences operating, albeit with variable efficiency, in different sequence contexts.

Although it is not known whether some or all of the selected sequences interact with splicing factors, it appears likely that they bind specific proteins which, in conjunction with splicing factors, form sequence-specific spliceosomes. A prime candidate for this model of exon recognition is the family of SR proteins (53; reviewed in reference 20). The individual SR proteins have been shown to have different effects on alternative splicing of pre-mRNAs with different sequences (54) and also depending on the presence of other proteins, such as hnRNPA1 (4, 28) and tra and tra2 (27). Recent results showed that SR proteins interact with exon enhancer sequences (26, 35, 39) and, depending on the sequence of the enhancer (35), form complexes that include different family members. In addition to the well characterized SR proteins (53), anti-SR monoclonal antibody detects a number of other proteins in HeLa cells (22a). Moreover, U1 and U2 small nuclear ribonucleoproteins (26, 41) and U2AF splicing factor (48) appear to interact with enhancer sequences, most likely via protein-protein interactions with SR protein factors. From these data, one can anticipate that numerous sequence-specific splicing complexes will be formed on exon sequences or that a large number of sequences may acquire a similar tertiary structure that is recognized by the splicing machinery. The diversity of sequences identified in this work may reflect these possibilities.

ACKNOWLEDGMENTS

We thank Elizabeth Smith for excellent technical assistance.
This work was supported in part by NIH grant GM32994 to R.K.

REFERENCES

1. **Amendt, B. A., D. Hesslein, L. J. Chang, and C. M. Stoltzfus.** 1994. Presence of negative and positive *cis*-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1. Mol. Cell. Biol. **14:**3960–3970.
2. **Black, D. L.** 1991. Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon on non-neuronal cells? Genes Dev. **5:**389–402.
3. **Black, D. L.** 1992. Activation of c-src neuron-specific splicing by an unusual RNA element in vivo and in vitro. Cell **69:**795–807.
4. **Caceres, J. F., S. Stamm, D. M. Helfman, and A. R. Krainer.** 1994. Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science **265:**1706–1709.
5. **Chen, I. T., and L. A. Chasin.** 1994. Large exon size does not limit splicing in vivo. Mol. Cell. Biol. **14:**2140–2146.
6. **Cooper, T. A.** 1992. In vitro splicing of cardiac troponin T precursors: exon mutations disrupt splicing of the upstream intron. J. Biol. Chem. **267:**5330–5338.
7. **Cooper, T. A., and C. P. Ordahl.** 1989. Nucleotide substitutions within the cardiac troponin T alternative exon disrupt pre-mRNA alternative splicing. Nucleic Acids Res. **17:**7905–7921.
8. **Cote, G. J., D. T. Stolow, S. Peleg, S. M. Berget, and R. F. Gagel.** 1992. Identification of exon sequences and an exon binding protein involved in alternative RNA splicing of calcitonin/CGRP. Nucleic Acids Res. **20:**2361–2366.
9. **Dirksen, W. P., R. K. Hampson, Q. Sun, and F. M. Rottman.** 1994. A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA. J. Biol. Chem. **269:**6431–6436.

10. **Dominski, Z., and R. Kole.** 1991. Selection of splice sites in pre-mRNA with short internal exons. Mol. Cell. Biol. **11:**6075–6083.
11. **Dominski, Z., and R. Kole.** 1992. Cooperation of pre-mRNA sequence elements in splice site selection. Mol. Cell. Biol. **12:**2108–2114.
12. **Dominski, Z., and R. Kole.** 1994. Identification of exon sequences involved in splice site selection. J. Biol. Chem. **269:**23590–23596.
13. **Fu, X.-D., R. A. Katz, A. M. Skalka, and T. Maniatis.** 1991. The role of branchpoint and 3′-exon sequences in the control of balanced splicing of avian retrovirus RNA. Genes Dev. **5:**211–220.
14. **Furdon, P. J., and R. Kole.** 1986. Inhibition of splicing but not cleavage at the 5′ splice site by truncating human beta-globin pre-mRNA. Proc. Natl. Acad. Sci. USA **83:**927–931.
15. **Furdon, P. J., and R. Kole.** 1988. The length of the downstream exon and the substitution of specific sequences affect pre-mRNA splicing in vitro. Mol. Cell. Biol. **8:**860–866.
16. **Green, R., A. D. Ellington, and J. W. Szostak.** 1990. In vitro genetic analysis of the Tetrahymena self-splicing intron. Nature (London) **347:**406–408.
17. **Hampson, R. K., L. La Follette, and F. Rottman.** 1989. Alternative processing of bovine growth hormone mRNA is influenced by downstream exon sequences. Mol. Cell. Biol. **9:**1604–1610.
18. **Hawkins, J. D.** 1988. A survey of intron and exon lengths. Nucleic Acids Res. **16:**9893–9908.
19. **Hedley, M. L., and T. Maniatis.** 1991. Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to tra-2 protein in vitro. Cell **65:**579–586.
20. **Horowitz, D. S., and A. R. Krainer.** 1994. Mechanisms for selecting 5′ splice sites in mammalian pre-mRNA splicing. Trends Genet. **10:**100–106.
20a. **Horwitz, M. S., and L. A. Loeb.** 1988. DNA sequences of random origin as probes of Escherichia coli promoter architecture. J. Biol. Chem. **263:**14724–14731.
21. **Inoue, K., K. Hoshijima, I. Higuchi, H. Sakamoto, and Y. Shimura.** 1992. Binding of the Drosophila transformer and transformer-2 proteins to the regulatory elements of doublesex primary transcript for sex-specific RNA processing. Proc. Natl. Acad. Sci. USA **88:**683–687.
22. **Katz, R. A., and A. M. Skalka.** 1990. Control of retroviral RNA splicing through maintenance of suboptimal processing signals. Mol. Cell. Biol. **10:**696–704.
22a. **Kole, R.** Unpublished data.
23. **Konarska, M. M., R. A. Padgett, and P. A. Sharp.** 1984. Recognition of cap structure in splicing in vitro of mRNA precursors. Cell **38:**731–736.
24. **Krainer, A. R., T. Maniatis, B. Ruskin, and M. R. Green.** 1984. Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell **36:**993–1005.
25. **Kuo, H.-C., F. H. Nassim, and P. J. Grabowski.** 1991. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. Science **251:**1045–1050.
26. **Lavigueur, A., H. La Branche, A. R. Kornblihtt, and B. Chabot.** 1993. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. Genes Dev. **7:**2405–2417.
27. **Lynch, K. W., and T. Maniatis.** 1995. Synergistic interactions between two distinct elements of a regulated splicing enhancer. Genes Dev. **9:**284–293.
28. **Mayeda, A., D. M. Helfman, and A. R. Krainer.** 1993. Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. Mol. Cell. Biol. **13:**2993–3001.
29. **Nagoshi, R. N., and B. S. Baker.** 1990. Regulation of sex-specific RNA splicing at the Drosophila doublesex gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. Genes Dev. **4:**89–97.
30. **Parent, A., S. Zeitlin, and A. Efstratiadis.** 1987. Minimal exon sequence splicing requirements for efficient in vitro splicing of monointronic nuclear pre-mRNA. J. Biol. Chem. **262:**11284–11291.
31. **Reed, R., and T. Maniatis.** 1986. A role for exon sequences and splice-site proximity in splice site selection. Cell **46:**681–690.
32. **Robberson, B. L., G. J. Cote, and S. M. Berget.** 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. **10:**84–94.
33. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
34. **Somaskhar, M. B., and J. E. Mertz.** 1985. Exon mutations that affect the choice of splice sites used in processing the SV40 late transcripts. Nucleic Acids Res. **13:**5591–5609.
35. **Staknis, D., and R. Reed.** 1994. SR-proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol. Cell. Biol. **14:**7670–7682.
36. **Sterner, D. A., and S. M. Berget.** 1993. In vivo recognition of a vertebrate mini-exon as an exon-intron-exon unit. Mol. Cell. Biol. **13:**2677–2687.
37. **Streuli, M., and H. Saito.** 1989. Regulation of tissue-specific alternative splicing: exon-specific cis-elements govern the splicing of leukocyte common antigen pre-mRNA. EMBO J. **8:**787–796.
38. **Sun, Q., R. K. Hampson, and F. M. Rottman.** 1993. In vitro analysis of bovine growth hormone pre-mRNA alternative splicing J. Biol. Chem. **268:**15659–15666.
39. **Sun, Q., A. Mayeda, R. K. Hampson, A. R. Krainer, and F. M. Rottman.** 1993. General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. Genes Dev. **7:**3519–3529.
40. **Talerico, M., and S. M. Berget.** 1990. Effect of 5′ splice site mutations on splicing of the preceding intron. Mol. Cell. Biol. **10:**6299–6305.
41. **Tanaka, K., A. Watakabe, and Y. Shimura.** 1994. Polypurine sequences within a downstream exon function as a splicing enhancer. Mol. Cell. Biol. **14:**1347–1354.
42. **Tian, M., and T. Maniatis.** 1992. A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. Cell **74:**105–114.
43. **Tian, M., and T. Maniatis.** 1992. Positive control of pre-mRNA splicing in vitro. Science **256:**237–240.
44. **Tian, M., and T. Maniatis.** 1994. A splicing enhancer exhibits both constitutive and regulated activities. Genes Dev. **8:**1703–1712.
45. **Tsai, A. Y. M., M. Streuli, and H. Saito.** 1989. Integrity of the exon 6 sequence is essential for tissue-specific alternative splicing of human leukocyte common antigen pre-mRNA. Mol. Cell. Biol. **9:**4550–4555.
46. **Tuerk, C., and L. Gold.** 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science **249:**505–510.
47. **Turnbull-Ross, A. D., A. J. Else, and I. C. Eperon.** 1988. The dependence of splicing efficiency on the length of 3′ exon. Nucleic Acids Res. **16:**395–411.
48. **Wang, Z., H. M. Hoffman, and P. J. Grabowski.** 1995. Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. RNA **1:**21–35.
49. **Wakamatsu, N., H. Kobayashi, T. Miyatake, and S. Tsuji.** 1992. A novel exon mutation in the human beta-hexosaminidase beta subunit gene affects 3′ splice site selection. J. Biol. Chem. **267:**2406–2413.
50. **Watakabe, A., H. Sakamoto, and Y. Shimura.** 1991. Repositioning of an alternative exon sequence of mouse IgM pre-mRNA activates splicing of the preceding intron. Gene Expression **1:**175–184.
51. **Watakabe, A., K. Tanaka, and Y. Shimura.** 1993. The role of exon sequences in splice site selection. Genes Dev. **7:**407–418.
52. **Xu, R., J. Teng, and T. Cooper.** 1993. The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. Mol. Cell. Biol. **13:**3660–3674.
53. **Zahler, A. M., S. W. Lane, J. A. Stolk, and M. B. Roth.** 1992. SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev. **6:**837–847.
54. **Zahler, A. M., K. M. Neugebauer, W. S. Lane, and M. B. Roth.** 1993. Distinct functions of SR proteins in alternative pre-mRNA splicing. Science **260:**219–222.